

# Data Viz Final - Mike Cuesta

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.2.1      v purrr  0.3.3
## v tibble  2.1.3      v dplyr  0.8.4
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library(ggplot2)
Heart_Stroke <- read_csv("data/Heart_Stroke.csv")

## Parsed with column specification:
## cols(
##   id = col_double(),
##   gender = col_character(),
##   age = col_double(),
##   hypertension = col_double(),
##   heart_disease = col_double(),
##   ever_married = col_character(),
##   work_type = col_character(),
##   Residence_type = col_character(),
##   avg_glucose_level = col_double(),
##   bmi = col_double(),
##   smoking_status = col_character(),
##   stroke = col_double()
## )
```

**Description:** The data set chosen comes from Kaggle. The data shows the information on hypertension, stroke, and heart disease cases. This is denoted 1; having these issues present, 0; issues are not present. An individual case can be identified by its id. We categorize cases by their gender, age, type of living location, and other factors such as smoking status, have they ever been married, the type of work they do, and if they are children. These along with things such as BMI and glucose levels help draw an image for possible stroke and heart issue concerns when analyzed properly.

## Graph 1

```
test1 <- ggplot(Heart_Stroke, aes(x=gender, y=hypertension, group=age, color=age))

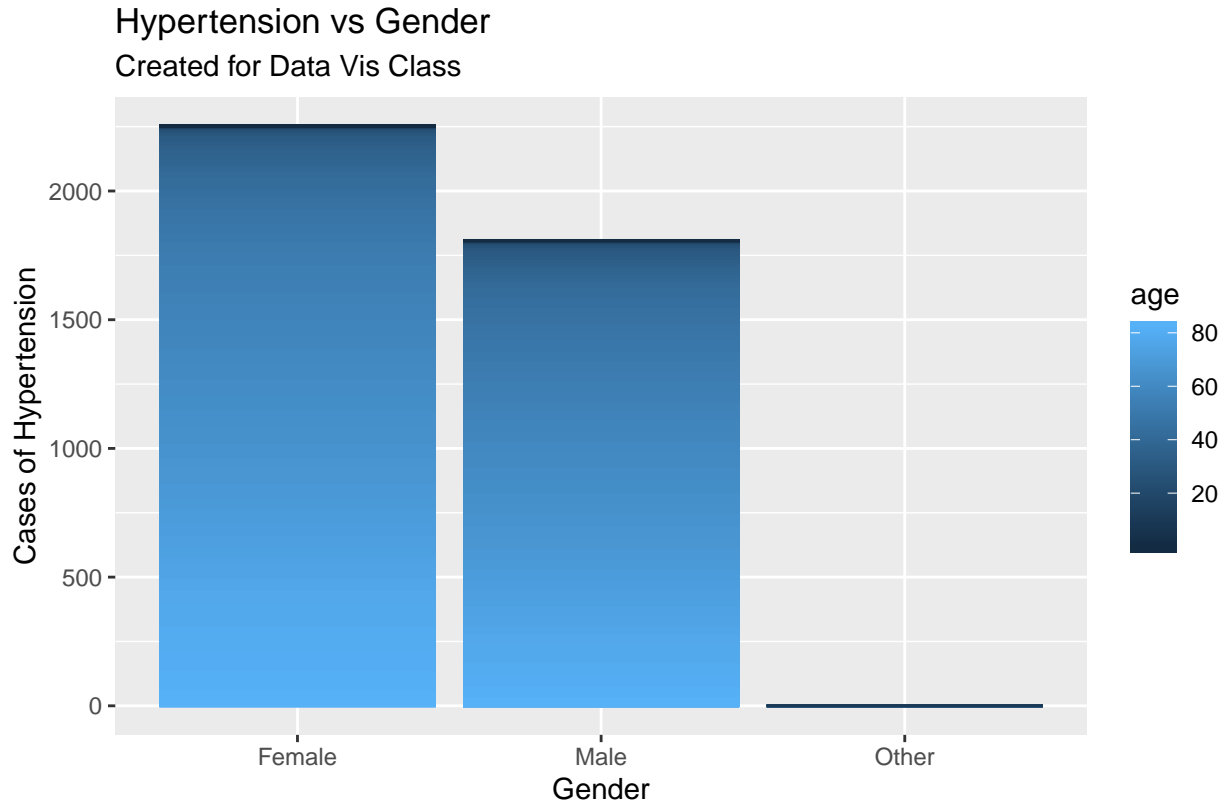
HYP_GENDER <- test1 + labs(x = "Gender",
                          y = "Cases of Hypertension",
```

```

title = "Hypertension vs Gender",
subtitle = "Created for Data Vis Class",
caption = "Source: https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data#train\_2"

```

```
HYP_GENDER + geom_col()
```



## Graph 2

```

test2 <- ggplot(Heart_Stroke, aes(x=age, y= hypertension))

test2a <- test2 + labs(x = "Age",
  y = "Cases of Hypertension",
  title = "Hypertension Cases Over Age",
  subtitle = "Created for Data Vis Class",
  caption = "Source: https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data#train\_2"

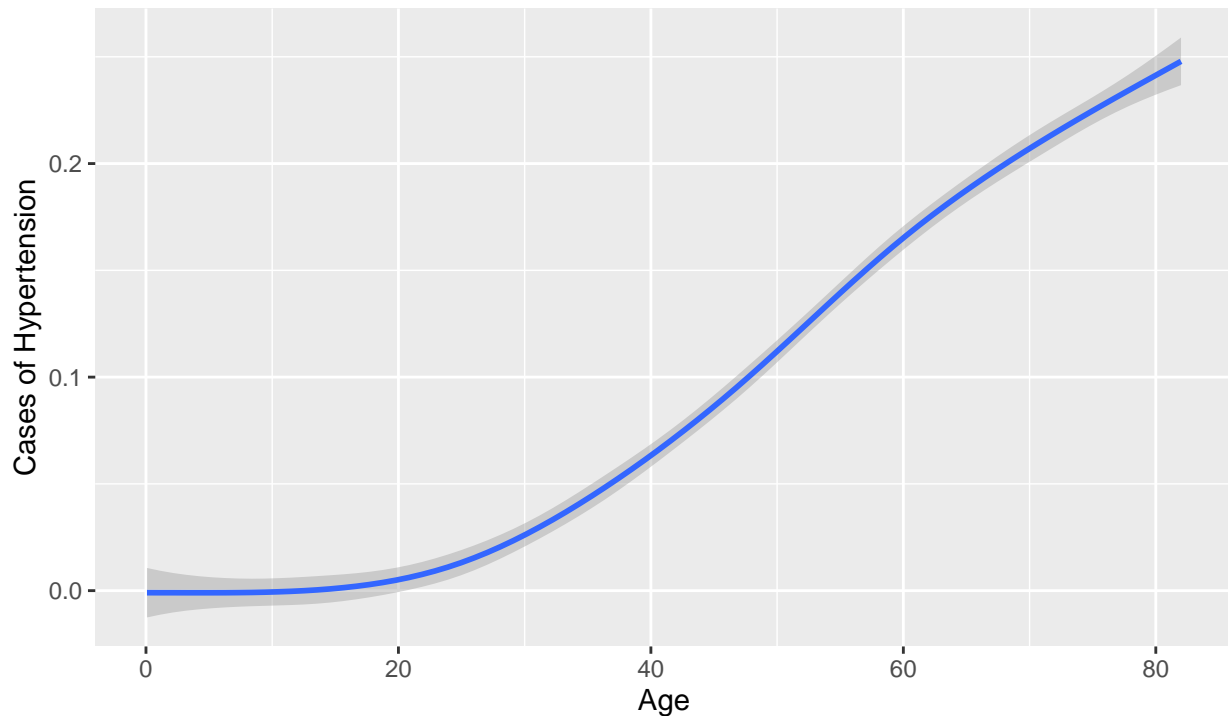
```

```
test2a + geom_smooth()
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

## Hypertension Cases Over Age

Created for Data Vis Class



Source: [https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data#train\\_2v.csv](https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data#train_2v.csv)

### Graph 3

```
test3 <- ggplot(Heart_Stroke, aes(x=age, y=hypertension, color=work_type))

HYP_AGE <- test3 + labs(x = "Age",
  y = "Cases of Hypertension",
  title = "Reports of Hypertension by Age Based off Type of Work",
  subtitle = "Created for Data Vis Class",
  caption = "Source: https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data#train\_2v.csv")

HYP_AGE + geom_bar(stat = "identity")
```

## Reports of Hypertension by Age Based off Type of Work

Created for Data Vis Class



Source: [https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data#train\\_2v.csv](https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data#train_2v.csv)

### Graph 4

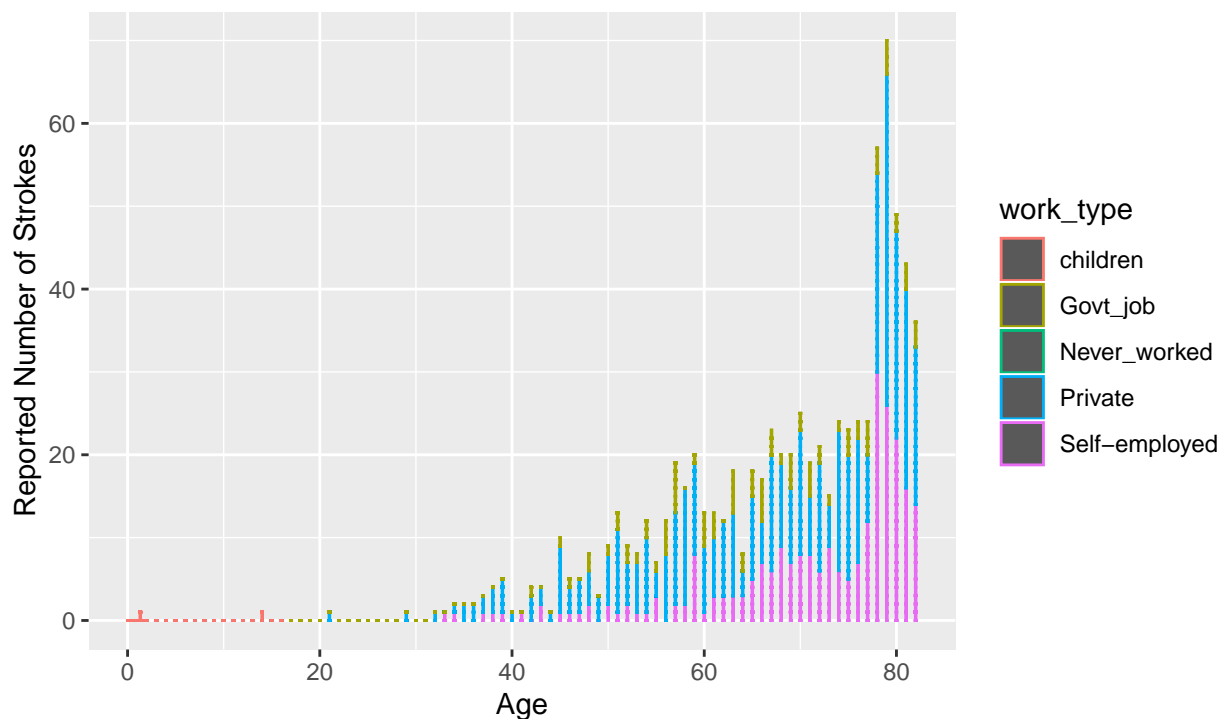
```
test4 <- ggplot(Heart_Stroke, aes(x=age, y=stroke, color=work_type))

STROKE_AGE <- test4 + labs(x = "Age",
  y = "Reported Number of Strokes",
  title = "Sum of Strokes by Age Based off Type of Work",
  subtitle = "Created for Data Vis Class",
  caption = "Source: https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data#train_2v.csv")

STROKE_AGE + geom_bar(stat = "identity")
```

## Sum of Strokes by Age Based off Type of Work

Created for Data Vis Class



Source: [https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data#train\\_2v.csv](https://www.kaggle.com/asaumya/healthcare-dataset-stroke-data#train_2v.csv)

## Graph 5

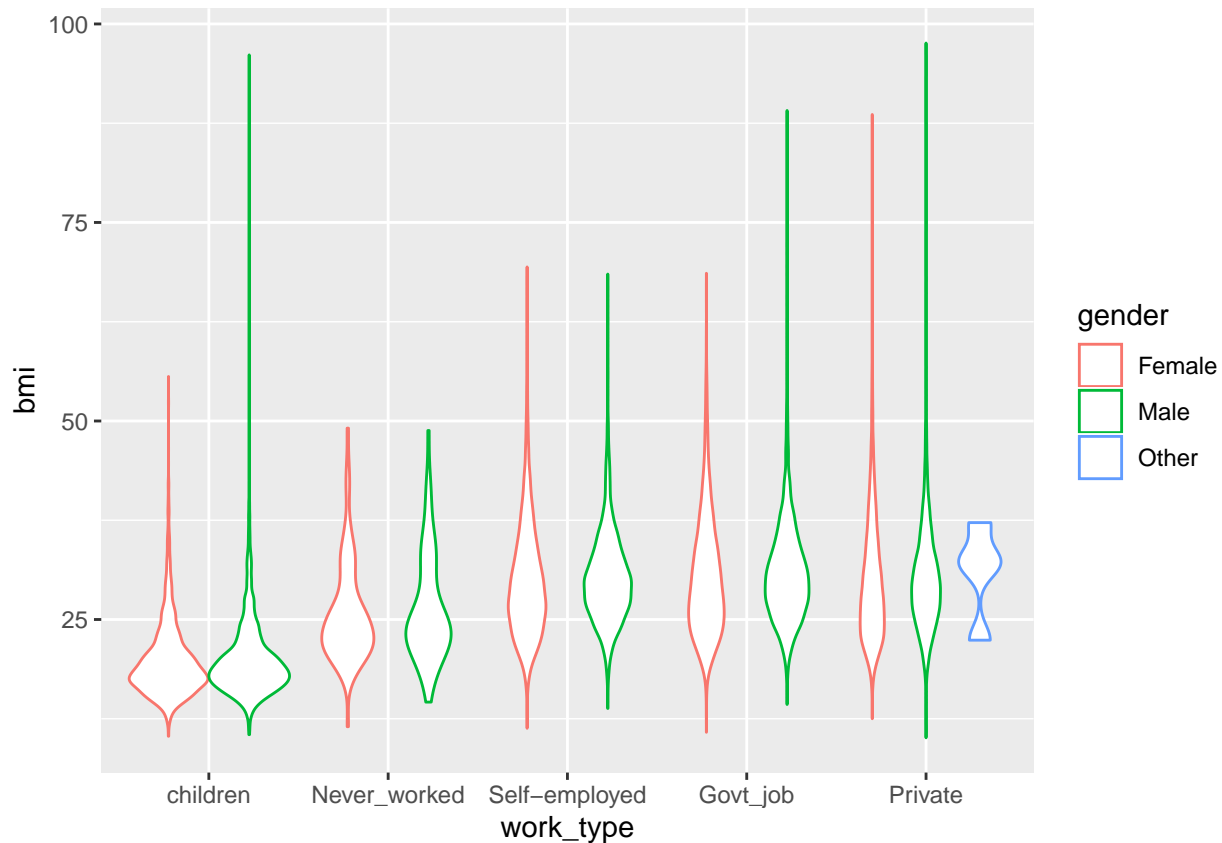
```
Heart_Stroke <-
  Heart_Stroke %>%
    mutate(work_type = fct_relevel(work_type, "children", "Never_worked", "Self-employed", "Govt_job", "Private"))

test5 <- ggplot(Heart_Stroke, aes(x=work_type, y=bmi, color=gender))

BMI_WORK <- test5

BMI_WORK + geom_violin()

## Warning: Removed 1462 rows containing non-finite values (stat_ydensity).
```



### Tableau Dashboard:

[https://public.tableau.com/profile/michael.cuesta#!/vizhome/Datavizfinal\\_15839652620660/HeartDiseaseoverLifeSpan?publish=yes](https://public.tableau.com/profile/michael.cuesta#!/vizhome/Datavizfinal_15839652620660/HeartDiseaseoverLifeSpan?publish=yes)

### Executive Summary:

Before creating these static graphics I spent time looking at the data on Tableau to see if I could find any trends. I wanted to look at men vs women and quickly found out there were a few more women in the study. However, it seemed that women who had never smoked had more cases of hypertension. After looking at gender I confirmed that as one ages there will be more cases of hypertension and strokes. I decided not to look at heart disease as an individual could have heart disease at any age but we only see high blood pressure or actual strokes in older individuals. Curious as to what could be causing this, I colored by work type and found something interesting. Individuals with private jobs or government jobs seem to both have the highest cases of stroke and hypertension when compared to other employment types. Lastly, I wondered if BMI had any influence but it appeared across the board that these values were pretty similar. This along with average glucose levels can be explored further on the dashboard, mainly by age. The data was easy to deal with. I did not do extra manipulation to the data as I felt the columns and values were comfortable to work with as is.